

# A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk

Matthias Heinig<sup>1,2\*</sup>, Enrico Petretto<sup>3,4\*</sup>, Chris Wallace<sup>5</sup>, Leonardo Bottolo<sup>3,4</sup>, Maxime Rotival<sup>6</sup>, Han Lu<sup>3</sup>, Yoyo Li<sup>3</sup>, Rizwan Sarwar<sup>3</sup>, Sarah R. Langley<sup>3</sup>, Anja Bauerfeind<sup>1</sup>, Oliver Hummel<sup>1</sup>, Young-Ae Lee<sup>1,7</sup>, Svetlana Paskas<sup>1</sup>, Carola Rintisch<sup>1</sup>, Kathrin Saar<sup>1</sup>, Jason Cooper<sup>3</sup>, Rachel Buchan<sup>3</sup>, Elizabeth E. Gray<sup>8</sup>, Jason G. Cyster<sup>8</sup>, Cardiogenics Consortium†, Jeanette Erdmann<sup>9</sup>, Christian Hengstenberg<sup>10</sup>, Serya Maouche<sup>6</sup>, Willem H. Ouwehand<sup>11,12</sup>, Catherine M. Rice<sup>12</sup>, Nilesh J. Samani<sup>13</sup>, Heribert Schunkert<sup>9</sup>, Alison H. Goodall<sup>13</sup>, Herbert Schulz<sup>1</sup>, Helge G. Roeder<sup>2</sup>, Martin Vingron<sup>2</sup>, Stefan Blankenberg<sup>14</sup>, Thomas Münzel<sup>14</sup>, Tanja Zeller<sup>14</sup>, Silke Szymczak<sup>15</sup>, Andreas Ziegler<sup>15</sup>, Laurence Tiret<sup>6</sup>, Deborah J. Smyth<sup>5</sup>, Michal Pravenec<sup>16</sup>, Timothy J. Aitman<sup>3</sup>, Francois Cambien<sup>6</sup>, David Clayton<sup>5</sup>, John A. Todd<sup>5</sup>, Norbert Hubner<sup>1,17</sup> & Stuart A. Cook<sup>3,18</sup>

**Combined analyses of gene networks and DNA sequence variation can provide new insights into the aetiology of common diseases that may not be apparent from genome-wide association studies alone. Recent advances in rat genomics are facilitating systems-genetics approaches<sup>1,2</sup>. Here we report the use of integrated genome-wide approaches across seven rat tissues to identify gene networks and the loci underlying their regulation. We defined an interferon regulatory factor 7 (IRF7<sup>3</sup>)-driven inflammatory network (IDIN) enriched for viral response genes, which represents a molecular biomarker for macrophages and which was regulated in multiple tissues by a locus on rat chromosome 15q25. We show that Epstein-Barr virus induced gene 2 (*Ebi2*, also known as *Gpr183*), which lies at this locus and controls B lymphocyte migration<sup>4,5</sup>, is expressed in macrophages and regulates the IDIN. The human orthologous locus on chromosome 13q32 controlled the human equivalent of the IDIN, which was conserved in monocytes. IDIN genes were more likely to associate with susceptibility to type 1 diabetes (T1D)—a macrophage-associated autoimmune disease—than randomly selected immune response genes ( $P = 8.85 \times 10^{-6}$ ). The human locus controlling the IDIN was associated with the risk of T1D at single nucleotide polymorphism rs9585056 ( $P = 7.0 \times 10^{-10}$ ; odds ratio, 1.15), which was one of five single nucleotide polymorphisms in this region associated with *EBI2* (*GPR183*) expression. These data implicate *IRF7* network genes and their regulatory locus in the pathogenesis of T1D.**

Although genome-wide association studies (GWASs) have uncovered many common genetic variants associated with human diseases, the molecular mechanisms by which DNA variation affects disease risk remain poorly characterized<sup>6</sup>. To translate genetic association into biological function, DNA variation has been correlated with gene expression to identify the genetic drivers of gene networks, which are coordinately regulated by transcription factors and represent important determinants of disease aetiology<sup>7–9</sup>. Here we used a panel of recombinant inbred rat strains<sup>1</sup> to study transcription-factor-driven gene networks and their regulatory loci and integrated these data with

human gene expression and GWAS data to identify genes, networks and pathways for human disease (Supplementary Fig. 1).

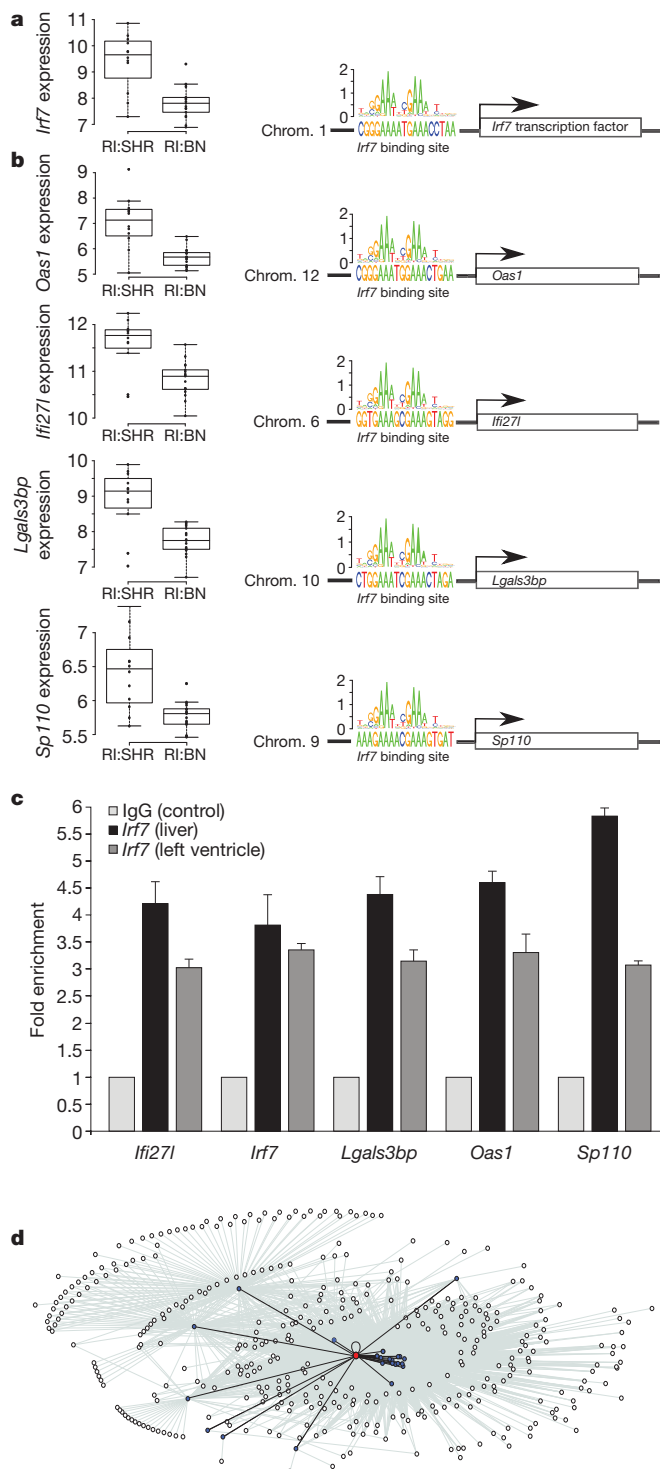
We combined expression quantitative trait loci (eQTLs) from fat, kidney and heart<sup>1,2</sup> with new eQTL data in aorta, skeletal muscle, adrenal gland and liver to create genome-wide eQTL data sets across seven rat tissues. We used a two-step procedure to integrate eQTLs and transcription factor target genes to identify transcription-factor-driven gene networks (Supplementary Information). In the first step, we identified 147 transcription factors whose expression mapped to 587 eQTLs across seven tissues, which were mostly (>90%) under *trans*-regulatory genetic control, in keeping with previous studies in yeast<sup>10,11</sup>. In the second step, we tested for enrichment of transcription factor binding sites (TFBSs)<sup>12</sup> in the putative promoter sequences of genes whose expression mapped to *trans*-eQTLs. Out of the 13 transcription-factor-driven gene networks identified (Supplementary Table 1) we observed the strongest TFBS enrichment for interferon regulatory transcription factor *Irf7* ( $P < 1 \times 10^{-6}$ ; false discovery rate (FDR),  $< 5 \times 10^{-5}$ ). *Irf7* TFBSs were predicted in the promoters of 23 genes, including *Irf7* itself, that all mapped to a single *trans*-eQTL on rat chromosome 15q25 in adrenal gland, kidney, heart and liver. We confirmed a subset of the predicted *Irf7* targets by chromatin immunoprecipitation and quantitative PCR that established direct interaction of *Irf7* with the promoters of these genes (Fig. 1a–c). Taken together, this provides evidence for a transcription-factor-driven regulatory cascade in which genetic variation on chromosome 15q25 modulates the expression of *Irf7* and *Irf7* target genes.

*Irf7* is a master regulator of the type 1 interferon response<sup>3</sup>, and genes directly regulated by *Irf7* may comprise the core components of a larger network, which we identified by genome-wide co-expression analysis of *Irf7* target genes across tissues (Supplementary Information). This revealed a network of 247 genes across seven tissues, which was expanded to 305 genes in four of the seven tissues where additional gene expression data were available (FDR < 0.1%) (Supplementary Table 2). Gene Ontology analysis of the network showed enrichment for specific biological processes, including ‘immune response’

<sup>1</sup>Max-Delbrück-Center for Molecular Medicine (MDC), Robert-Rössle-Straße 10, 13125 Berlin, Germany. <sup>2</sup>Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. <sup>3</sup>Medical Research Council Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK. <sup>4</sup>Department of Epidemiology and Biostatistics, Faculty of Medicine, Imperial College London, Praed Street, London W2 1PG, UK. <sup>5</sup>Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. <sup>6</sup>INSERM UMR5 937, Pierre and Marie Curie University (UPMC, Paris 6) and Medical School, 91 Boulevard de l'Hôpital, Paris 75013, France. <sup>7</sup>Pediatric Pneumology and Immunology, Charité - Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. <sup>8</sup>Howard Hughes Medical Institute and Department of Microbiology and Immunology, University of California San Francisco, California 94143, USA. <sup>9</sup>Universität zu Lüneburg, Medizinische Klinik II, 23538 Lüneburg, Germany. <sup>10</sup>Klinik und Poliklinik für Innere Medizin II, Universität Regensburg, 93053 Regensburg, Germany. <sup>11</sup>Department of Haematology, University of Cambridge and National Health Service Blood and Transplant, Cambridge CB2 0PT, UK. <sup>12</sup>Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton CB10 1SA, UK. <sup>13</sup>Department of Cardiovascular Sciences, University of Leicester and Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester LE3 9QP, UK. <sup>14</sup>Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Universitätsmedizin, Langenbeckstrasse 1, 55131 Mainz, Germany. <sup>15</sup>Institut für Medizinische Biometrie und Statistik, Universität zu Lüneburg, Universitätsklinikum Schleswig-Holstein, Campus Lüneburg, Maria-Goeppert-Straße 1, 23562 Lüneburg, Germany. <sup>16</sup>Institute of Physiology, Czech Academy of Sciences and Centre for Applied Genomics, Videnska 1083, 14220 Prague 4, Czech Republic. <sup>17</sup>CC4, Campus Charité Mitte, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. <sup>18</sup>National Heart and Lung Institute, Imperial College, Dovehouse Street, London SW3 6LY, UK.

\*These authors contributed equally to this work.

†A list of participants and their affiliations appears at the end of the paper.



( $P = 7.5 \times 10^{-29}$ ), and 'response to virus' ( $P = 9.6 \times 10^{-7}$ ) (Supplementary Table 3). We designated the network the *Irf7*-driven inflammatory gene network (IDIN) (Fig. 1d), which was most enriched for expression in mouse bone marrow macrophages ( $P = 1.6 \times 10^{-159}$ ) and human monocytes ( $P = 6.0 \times 10^{-177}$ ), with high levels of expression in other immune cells, including B lymphocytes (Supplementary Fig. 2).

Although a core of 23 *Irf7* target genes mapped to the same *trans*-eQTL on rat chromosome 15, the overall genetic control of the IDIN was unknown. We used sparse Bayesian regression models<sup>13</sup> to determine the association between expression levels of IDIN genes across seven tissues with genome-wide single nucleotide polymorphisms

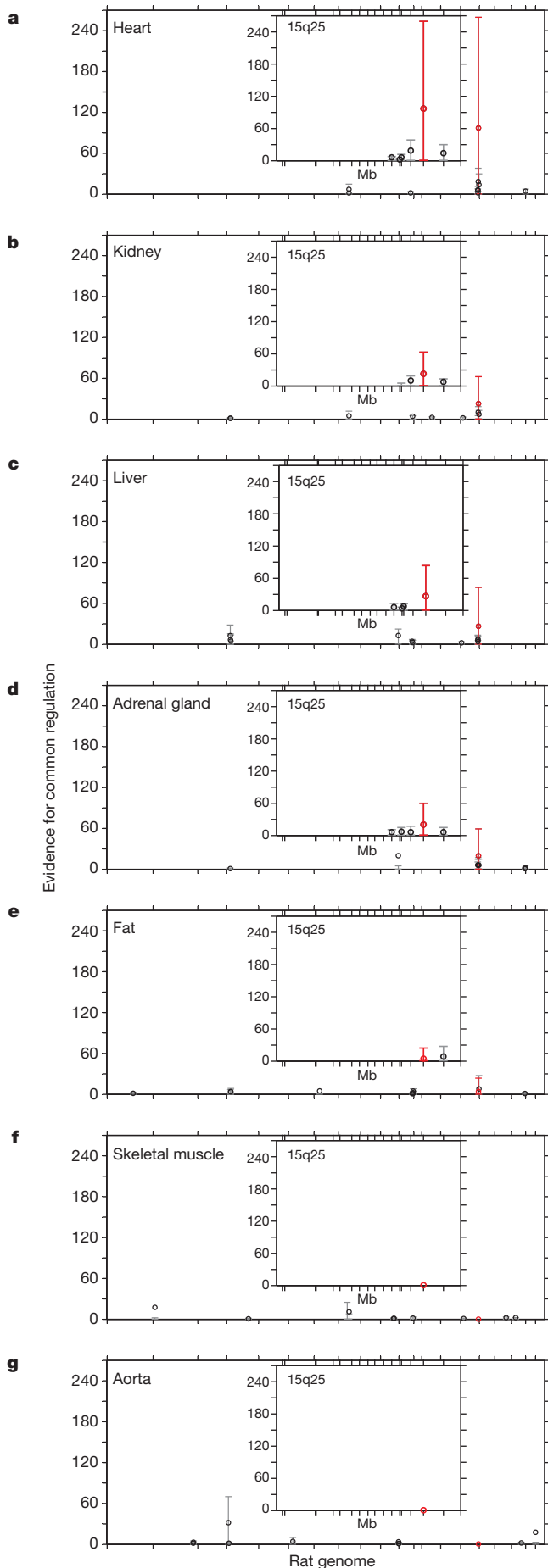
**Figure 1 | The rat *Irf7*-driven inflammatory gene network.** a, b, *Trans*-regulated expression of *Irf7* (a) and genes containing *Irf7* TFBSs (b) by rat chromosome 15q25. Left panels, gene expression in the left ventricle is shown in the recombinant inbred rat strains grouped by SHR or Brown Norway genotype at SNP J666808 (SHR allele, RI:SHR; Brown Norway allele, RI:BN). Data are presented as box plots (box, 25th–75th percentiles; solid bar, median; whiskers, 10th–90th percentiles; total  $n = 29$ ). Right panels, TFBS predictions are represented for the five (out of 23 predicted) *Irf7* target genes. The chromosome encoding the *Irf7* target is shown to the left of the predicted *Irf7* binding sites. These data provide evidence for a regulatory cascade in which a locus on chromosome 15q25 regulates the expression of *Irf7* on chromosome 1 in an allele-dependent manner with consequent effects on *Irf7* target genes mediated through *Irf7* TFBSs. c, Quantitative chromatin immunoprecipitation of predicted *Irf7* target genes. Direct binding of *Irf7* to the promoters of the predicted targets *Ifi271* (*Ifi27*), *Irf7*, *Lgals3bp*, *Oas1* (*Oas1a*) and *Sp110* was confirmed in liver and heart tissues. Fold enrichments are shown relative to non-immune immunoglobulin-G (IgG) control. Error bars, s.d. ( $n = 5$ ). d, The expanded IDIN comprising 305 genes. Nodes represent genes; the node representing *Irf7* is coloured red and its predicted targets are coloured blue (Supplementary Table 2). Edges connect genes that are either predicted *Irf7* targets (black) or show significant Pearson correlation (FDR < 0.1%) to one of the predicted targets (grey).

(SNPs) and identify regulatory 'hot spots'<sup>14</sup>. The same rat 15q25 locus, which controlled *Irf7* and its targets *in trans*, was associated with IDIN expression in all tissues (FDR < 1%) and showed the strongest evidence for common regulation in five out of seven tissues with increased expression of IDIN genes associated with the spontaneously hypertensive rat (SHR) allele (Fig. 2). The IDIN, which is highly expressed in immune cells, may represent a molecular signature of macrophages that are associated with risk of common inflammatory diseases<sup>15</sup> and autoimmune disease T1D<sup>16</sup>. Hence, we characterized expression of *Cd68*, an established marker of macrophages<sup>17</sup>, in SHR and Brown Norway hearts and the recombinant inbred strains. *Cd68* messenger RNA levels were elevated in SHR relative to Brown Norway heart ( $P = 0.01$ ), which reflected increased numbers of macrophages ( $P = 2 \times 10^{-22}$ ). In the recombinant inbred strains, *Cd68* was under *trans*-acting genetic control at the 15q25 locus that regulates the IDIN (Supplementary Fig. 3).

We then analysed genetic variation in the recombinant inbred strains using SNPs<sup>18</sup> from the 15q25 region, which contains seven annotated protein-coding genes, and determined the expression of IDIN genes in seven inbred rat strains of known genotype that refined the locus to a 700-kilobase region (Supplementary Fig. 4). Using the SHR genome sequence<sup>19</sup>, only *Dock9*, *Ebi2* and *Tm9sf2* showed DNA variation within the region, which was synonymous for *Dock9*, non-synonymous but not predicted to be functional for *Tm9sf2*, and a 5' untranslated region SNP for *Ebi2* (Supplementary Table 4). *Ebi2* was the only differentially expressed gene between parental strains within the region and was *cis*-regulated in heart and kidney and highly expressed in myeloid cell types (Supplementary Figs 4 and 5). We assessed the effect of the *Ebi2* 5' untranslated region SNP by luciferase assay; the SHR allele resulted in reduced luciferase activity relative to the Brown Norway allele (Supplementary Fig. 5).

*Ebi2* encodes an orphan G-protein-coupled receptor that controls B-cell migration<sup>4,5</sup> and is a candidate for the regulation of the IDIN at the chromosome 15q25 region. We localized *Ebi2* expression to *Cd68*<sup>+</sup> macrophages within the rat heart (Supplementary Fig. 6), an observation that we confirmed and extended across tissues (pancreas, liver, kidney and heart) in the *Ebi2*<sup>GFP/+</sup> mouse<sup>4</sup> (Supplementary Fig. 7). Short interfering RNA knockdown of *Ebi2* in primary cultures of rat macrophages (Supplementary Fig. 8a) increased expression of *Irf7*, the central hub of the IDIN, and of IDIN genes (Supplementary Fig. 8b). This suggests that *Ebi2* is a negative regulator of the innate immune response in macrophages, which would be consistent with lower *Ebi2* expression in the SHR, which has more macrophages than the Brown Norway rat (Supplementary Fig. 3).

To translate our findings to humans, we tested whether the IDIN was recapitulated in human immune cells using genome-wide expression



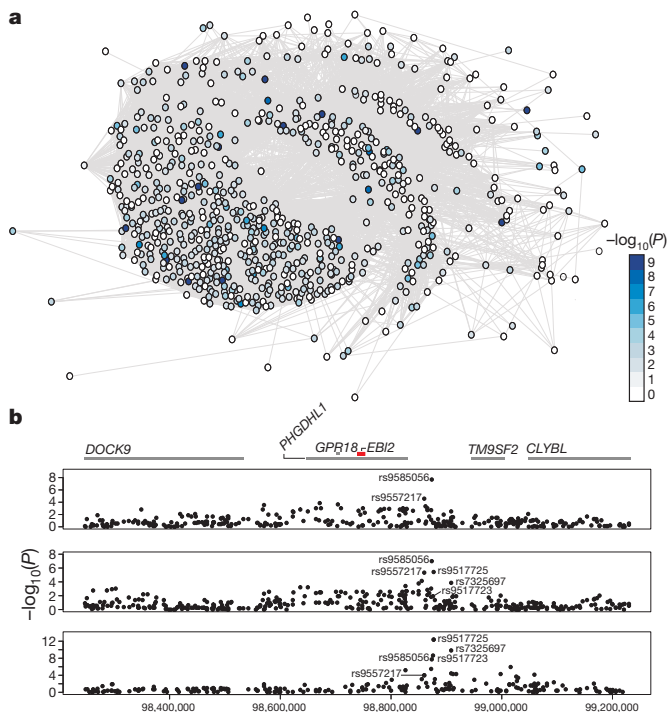
**Figure 2 | Genetic mapping of regulatory hot spots for the IDIN.** a–g, For each rat autosomal chromosome (horizontal axes), the strength of evidence for a SNP being a regulatory hot spot for controlling the network expression in seven tissues is measured by the average Bayes factor (vertical axes). Controlling the FDR at the 1% level for each eQTL, the average Bayes factor indicates the evidence in favour of common genetic regulation versus no genetic control, and is reported as a ratio between the strengths of these models (Supplementary Information). For the ten largest regulatory hot spots, the average Bayes factors (circles) and their 90% range (5th–95th percentiles, bars) are reported; a single SNP (J666808) that is consistently and most strongly associated with the network in five out of seven tissues is highlighted in red. Insets, average Bayes factors and 90% range for the SNPs on rat chromosome 15q25 (base pairs 87,479,238 to 108,949,015). SNP positions in the region are indicated by tick marks. Mb, megabase.

data from monocytes isolated from 1,490 individuals from the Gutenberg Heart Study<sup>20</sup> (GHS). We performed TFBS enrichment and co-expression analysis, analogous to that performed in the rat, and identified the human *IRF7*-driven network (Supplementary Table 5), which had strong overlap with the rat IDIN ( $P = 9.1 \times 10^{-20}$ ) and was most significantly annotated by the Gene Ontology term ‘response to virus’ ( $P = 1.9 \times 10^{-13}$ ) (Supplementary Table 6). Using monocyte gene expression data from a distinct cohort of 758 subjects from the Cardiogenics Study (Supplementary Information), we found the same set of co-regulated *IRF7* target genes (Supplementary Table 5) and significant overlap with the expanded *IRF7*-driven network identified in the GHS ( $P = 8.3 \times 10^{-23}$ ).

We determined whether the human chromosome 13q32 locus (spanning  $\sim 1$  Mb; Supplementary Table 7), which is orthologous to the critical rat chromosome 15q25 region, was associated with expression of IDIN genes in humans. Multivariate analysis of the Cardiogenics Study monocyte expression and genotype data revealed that six SNPs in the 13q32 region (including rs9557217 ( $P = 5.0 \times 10^{-5}$ ) and rs9585056 ( $P = 1.1 \times 10^{-3}$ )) were associated with *trans*-regulated expression of *IRF7* and *IRF7* target genes (Supplementary Fig. 9). We did not, however, detect a signal for *trans*-regulation of *IRF7* or *IRF7* target genes at the 13q32 locus in the GHS cohort. This may reflect differences between the monocyte selection protocols used in the two studies (Supplementary Information and data not shown).

In both the GHS and Cardiogenics Study cohorts, *EBI2* expression in monocytes was *cis*-regulated at the 13q32 locus, but the peak SNPs differed between the two cohorts (most-associated SNPs: Cardiogenics Study, rs9585056 ( $P = 2.2 \times 10^{-8}$ ); GHS, rs9517725 ( $P = 6.8 \times 10^{-13}$ )) (Fig. 3). However, a formal hypothesis test<sup>21</sup> of a common causal genetic variant was not rejected ( $P = 0.14$ ). Two of the five SNPs contained in the model explaining *EBI2* expression, rs9557217 and rs9585056, also had a significant *trans*-effect on IDIN gene expression in the Cardiogenics Study cohort (Supplementary Fig. 9), suggesting common regulatory control by this locus on the *IRF7* network and *EBI2* expression.

Monocyte-derived macrophages are critical determinants of inflammatory processes important for common diseases<sup>15</sup>, including autoimmune T1D<sup>22</sup>. The IDIN is expressed in macrophages, enriched for immune response genes, and contained *IFIH1*, a well-characterized T1D susceptibility gene<sup>23,24</sup>. We evaluated the association of the human orthologues of rat IDIN genes and genes in the human IDIN (Fig. 3) with T1D (Supplementary Information). SNPs close to ( $\leq 1$  Mb from) any IDIN genes were significantly more likely to associate with T1D in large-scale GWASs than SNPs close to genes not in the network ( $P = 2.4 \times 10^{-10}$ ) (Supplementary Table 8). We also tested the IDIN association with T1D against all genes annotated by the Gene Ontology term ‘immune response’ and established an over-representation of T1D-associated genes ( $P = 8.85 \times 10^{-6}$ ), indicating that the IDIN more specifically categorizes T1D genes than the Gene Ontology term ‘immune response’. The association of the IDIN with T1D genes remained when the human leukocyte antigen locus was removed from the analysis ( $P = 8.57 \times 10^{-4}$ ; Supplementary Table 8).



**Figure 3 | A gene network and locus for T1D risk.** **a**, Schematic of the union of *IRF7*-driven gene networks ( $n = 697$ ) that was created using the set of human orthologues of rat IDIN genes ( $n = 252$ ) and human IDIN genes ( $n = 495$ ). A Wilcoxon rank test showed SNPs close to IDIN genes to be significantly more likely to associate with T1D in large-scale GWASs than SNPs close to randomly selected genes ( $P = 2.5 \times 10^{-10}$ ) and randomly selected immune response genes ( $P = 8.8 \times 10^{-6}$ ). Nodes represent IDIN genes and the node colour indicates the  $P$  value (negative log scale) of the association of SNPs within 1 Mb of any IDIN gene with T1D (see Methods and <http://www.t1dbase.org>). **b**, Results of T1D association (top) and *EB12* eQTL analysis in the Cardiogenics Study (middle) and of *EB12* eQTL analysis in the GHS (bottom) at the human chromosome 13q32 locus that is orthologous to the 700-kilobase rat chromosome 15q25 region. The top panel shows the  $-\log_{10}(P)$  values of T1D association with SNPs in the region. SNP rs9585056 showed the strongest association with T1D ( $P = 7.0 \times 10^{-10}$ ) among the genotyped markers. The middle and bottom panels show the nominal  $-\log_{10}(P)$  values of marker regression against gene expression of *EB12* for all SNPs in the region. We defined *EB12* eQTL models by selecting SNPs using lasso regression (Supplementary Information) in the GHS (rs9585056, rs9517723, rs7325697). When adding imputed SNPs, rs9517725 explains most of the variation of the *EB12* expression ( $P = 6.8 \times 10^{-13}$ ) at this locus. Lasso model selection in the Cardiogenics Study yielded an overlapping set of three SNPs (rs9557217, rs9585056, rs9517723). The locations of genes in the region are depicted on the upper horizontal axis. On the lower horizontal axis, the genomic position on chromosome 13 is shown in base pairs.

In a GWAS meta-analysis of T1D in 7,514 cases and 9,045 controls<sup>25</sup>, we found evidence for association of the chromosome 13q32 region at SNP rs9585056 ( $P = 1.3 \times 10^{-7}$ ) that had not been reported before (Fig. 3b). We genotyped this SNP in two independent large cohorts and increased the strength of the T1D association (combined  $P = 7.0 \times 10^{-10}$ ; odds ratio (95% confidence interval), 1.15 (1.09–1.21); Supplementary Table 9). The minor C allele of SNP rs9585056 was associated with T1D risk, lower *EB12* expression in both GHS and Cardiogenics Study cohorts, and, on average, increased expression levels of IDIN genes in the Cardiogenics Study cohort. Although we cannot discriminate between single and multiple causal variants, overall these results show an overlap of association signals in the same region on human chromosome 13q32 for IDIN genes, *EB12* cis-regulation and T1D. We also noted that the *EB11* (*CCR7*) and *EB13* genes are also associated with T1D susceptibility: *EB11* is in the confirmed T1D region 17q21.2<sup>25</sup>, and *EB13* encodes the  $\beta$ -subunit of the IL-27

cytokine, for which the  $\alpha$ -subunit gene, *IL27*, is in the T1D region 16p11.2<sup>25</sup>, suggesting a link between Epstein–Barr virus infection and T1D.

The immunopathology of autoimmune T1D is characterized by infiltration of the pancreas with B and T lymphocytes and macrophages<sup>16</sup>. We have shown that IDIN genes contribute to T1D risk and implicate the innate viral response pathway and macrophages in the aetiology of T1D. Loci that perturb gene networks can be important for disease risk<sup>8</sup> and the new T1D susceptibility locus that we identified may regulate innate immune response genes in macrophages, as we demonstrated in the rat. *Ebi2*, which controls *Irf7*<sup>3</sup>, represents a candidate for *trans*-regulation of the human IDIN and for T1D risk. A role for *IRF7* in the pathogenesis of T1D is supported by functional studies<sup>26</sup> and by other T1D genes, namely *TLR7*, *TLR8*<sup>27</sup> and *IFIH1*<sup>23,24</sup>, which are regulated by or act through *IRF7*<sup>28</sup>. Our study shows that co-expression networks across species provide functional annotation of genes in biological processes that can be used to reveal the signal of common genetic variation of small effect that is not detected by GWASs.

## METHODS SUMMARY

We generated genome-wide expression data in the rat from seven tissues (adrenal gland, aorta, fat, kidney, left ventricle, liver and skeletal muscle) using Affymetrix RAE 230a and RAE 230\_2 chips. eQTL mapping was carried out using the genetic map of the BXH/HXB recombinant inbred strains generated in a previous large-scale effort by the STAR consortium<sup>18</sup>, as previously described<sup>1,2</sup>. In humans, expression data from isolated monocytes were obtained from 1,490 population-based individuals from the GHS<sup>20</sup> and from 758 individuals from the Cardiogenics Study. eQTL data were analysed in conjunction with TFBS enrichment analysis using PASTAA<sup>12</sup> to identify core gene networks centred on transcription factors. The core networks were expanded to include genes showing co-expression ( $FDR < 0.1\%$ ) with any of the core network genes in seven rat tissues and isolated human monocytes. We determined association between expression levels of the network genes and genome-wide SNPs in the rat using sparse Bayesian regression models<sup>13</sup>, and identified the major regulatory control points (hot spots)<sup>14</sup> for the entire network. Genes at the locus associated with the rat network were characterized by DNA sequencing, RNA sequencing, quantitative PCR analyses, luciferase assay and combined *in situ* hybridization and immunohistochemistry. A combined network, comprising the union or intersection of the rat and human networks, was constructed and analysed for association with T1D by means of a stratified Wilcoxon rank test to compare SNPs genotyped in T1D GWASs<sup>25,29</sup> close to ( $\leq 1$  Mb from) any network gene or to those close to any gene not in the network (see <http://www.t1dbase.org> for all T1D SNP association data). SNPs across the human locus, that is, orthologous to rat chromosome 15q25 controlling the network, were tested for association with T1D as described elsewhere<sup>25</sup>. Supplementary Fig. 1 provides an overview of the study design. Full methods are provided in Supplementary Information.

Received 18 December 2009; accepted 28 July 2010.

Published online 8 September 2010.

- Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genet.* **37**, 243–253 (2005).
- Petretto, E. *et al.* Integrated genomic approaches implicate osteoglycin (*Ogn*) in the regulation of left ventricular mass. *Nature Genet.* **40**, 546–552 (2008).
- Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
- Pereira, J. P., Kelly, L. M., Xu, Y. & Cyster, J. G. *EB12* mediates B cell segregation between the outer and centre follicle. *Nature* **460**, 1122–1126 (2009).
- Gatto, D., Paus, D., Basten, A., Mackay, C. R. & Brink, R. Guidance of B cells by the orphan G protein-coupled receptor *EB12* shapes humoral immune responses. *Immunity* **31**, 259–269 (2009).
- Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
- Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
- Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
- Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
- Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Yvert, G. *et al.* *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35**, 57–64 (2003).

12. Roeder, H. G., Manke, T., O'Keefe, S., Vingron, M. & Haas, S. A. PASTAA: identifying transcription factors associated with sets of co-regulated genes. *Bioinformatics* **25**, 435–442 (2009).
13. Petretto, E. *et al.* New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Comput. Biol.* **6**, e1000737 (2010).
14. Breittling, R. *et al.* Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4**, e1000232 (2008).
15. Nathan, C. & Ding, A. Nonresolving inflammation. *Cell* **140**, 871–882 (2010).
16. Eizirik, D. L., Colli, M. L. & Ortis, F. The role of inflammation in insulinitis and  $\beta$ -cell loss in type 1 diabetes. *Nature Rev. Endocrinol.* **5**, 219–226 (2009).
17. Holness, C. L. & Simmons, D. L. Molecular cloning of CD68, a human macrophage marker related to lysosomal glycoproteins. *Blood* **81**, 1607–1613 (1993).
18. Saar, K. *et al.* SNP and haplotype mapping for genetic analysis in the rat. *Nature Genet.* **40**, 560–566 (2008).
19. Atanur, S. S. *et al.* The genome sequence of the spontaneously hypertensive rat: analysis and functional significance. *Genome Res.* **20**, 791–803 (2010).
20. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
21. Plagnol, V., Smyth, D. J., Todd, J. A. & Clayton, D. G. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics* **10**, 327–334 (2009).
22. von Herrath, M. Diabetes: a virus–gene collaboration. *Nature* **459**, 518–519 (2009).
23. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
24. Smyth, D. J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. *Nature Genet.* **38**, 617–619 (2006).
25. Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genet.* **41**, 703–707 (2009).
26. Li, Q. *et al.* Interferon- $\alpha$  initiates type 1 diabetes in nonobese diabetic mice. *Proc. Natl Acad. Sci. USA* **105**, 12439–12444 (2008).
27. Cooper, J. D. *et al.* Follow-up of 1715 SNPs from the Wellcome Trust Case Control Consortium genome-wide association study in type I diabetes families. *Genes Immun.* **10** (suppl. 1), S85–S94 (2009).
28. Kawai, T. *et al.* Interferon- $\alpha$  induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
29. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge funding from the German National Genome Research Network (NGFN-Plus 'Genetics of Heart Failure'), the Helmholtz Association Alliance on Systems Biology (MSBN), EURATools (LSHG-CT-2005-019015), European Union FP6 (LSHM-CT-2006-037593), PHC ALLIANCE 2009 (19419PH), UK National Institute for Health Research Biomedical Research Unit (Royal Brompton and Harefield NHS Trusts, University Hospitals of Leicester NHS Trusts) and Biomedical Research Centre (Imperial College NHS Trust) awards, the British Heart Foundation, grant P301/10/0290 from the Grant Agency of the Czech Republic, grant 1M6837805002 from the Ministry of Education of the Czech Republic, the Fondation Leducq, the Medical Research Council UK, Research Councils UK, the Juvenile Diabetes Research Foundation International, National Institute for Health Research (UK), National Institute of Diabetes and Digestive and Kidney Diseases (USA), and the Wellcome Trust. The

research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. HEALTH-F4-2010-241504 (EURATRANS). O. Burren performed T1DBase analyses.

**Author Contributions** S.A.C., N.H. and E.P. initiated the study. M.H., E.P., N.H. and S.A.C. participated in the conception, design and coordination of the study. H.L., Y.L., R.S., Y.A.L., S.P., C.R., K.S. and R.B. performed genetic, biochemical and functional analyses in rats. E.E.G. and J.G.C. provided *Ebi2<sup>GFP/+</sup>* mouse data. M.P. and T.J.A. contributed materials and discussion of the manuscript. M.H., E.P., C.W., D.J.S., D.C., A.B., S.R.L., L.B., M.R. and L.T. designed and applied the modelling methodology and statistical analyses. M.H., E.P. and H. Schulz performed eQTL analysis in the rat. L.B. designed and performed the Bayesian analysis. C.W., D.J.S. and D.C. performed association analyses in humans. M.H., O.H., H.R. and M.V. designed and performed bioinformatics analyses in rats. J.E., C.H., S.M., W.H.O., C.M.R., N.J.S., H. Schunkert, A.H.G., S.B., T.M., T.Z., S.S., A.Z., M.R., L.T. and F.C. provided the human monocyte expression data and contributed to the transcriptomic analyses in the Cardiogenics Study and Gutenberg Heart Study cohorts. M.H., E.P., N.H. and S.A.C. wrote the paper with significant contributions from C.W. and J.A.T. All authors discussed the results and commented on the manuscript.

**Author Information** Microarray expression data in the rat have been deposited at ArrayExpress with the following identity codes: skeletal muscle, E-TABM-458; aorta, E-MTAB-322; liver, E-MTAB-323. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.H. ([nhuebner@mdc-berlin.de](mailto:nhuebner@mdc-berlin.de)) or S.A.C. ([stuart.cook@csc.mrc.ac.uk](mailto:stuart.cook@csc.mrc.ac.uk)).

#### Cardiogenics Consortium

Peter Braund<sup>1</sup>, Jay Gracey<sup>1</sup>, Unni Krishnan<sup>1</sup>, Jasbir S. Moore<sup>1</sup>, Chris P. Nelson<sup>1</sup>, Helen Pollard<sup>1</sup>, Tony Attwood<sup>2</sup>, Abi Crisp-Hihn<sup>2</sup>, Nicola Foad<sup>2</sup>, Jennifer Jolley<sup>2</sup>, Heather Lloyd-Jones<sup>2</sup>, David Muir<sup>2</sup>, Elizabeth Murray<sup>2</sup>, Karen O'Leary<sup>2</sup>, Angela Rankin<sup>2</sup>, Jennifer Sambrook<sup>2</sup>, Tiphaine Godfroy<sup>3</sup>, Jessy Brocheton<sup>3</sup>, Carole Proust<sup>3</sup>, Gerd Schmitz<sup>4</sup>, Susanne Heimerl<sup>5</sup>, Ingrid Lugauer<sup>5</sup>, Stephanie Belz<sup>6</sup>, Stefanie Gulde<sup>6</sup>, Patrick Linsel-Nitschke<sup>6</sup>, Hendrik Sager<sup>6</sup>, Laura Schroeder<sup>6</sup>, Per Lundmark<sup>7</sup>, Ann-Christine Syvannen<sup>7</sup>, Jessica Neudert<sup>8</sup>, Michael Scholz<sup>8</sup>, Panos Deloukas<sup>9</sup>, Emma Gray<sup>9</sup>, Rhian Williams<sup>9</sup> & David Niblett<sup>9</sup>.

<sup>1</sup>Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. <sup>2</sup>Department of Haematology, University of Cambridge and National Health Service Blood and Transplant, Cambridge CB2 2PT, UK. <sup>3</sup>INSERM UMRS 937, Pierre and Marie Curie University (UPMC, Paris 6) and Medical School, 91 Boulevard de l'Hôpital, Paris 75013, France. <sup>4</sup>Institut für Klinische Chemie und Laboratoriumsmedizin, Universität Regensburg, 93053 Regensburg, Germany. <sup>5</sup>Klinik und Poliklinik für Innere Medizin II, Universität Regensburg, 93053 Regensburg, Germany. <sup>6</sup>Universität zu Lübeck, Medizinische Klinik II, 23538 Lübeck, Germany. <sup>7</sup>Molecular Medicine, Department of Medical Sciences, Uppsala University, SE-751 85 Uppsala, Sweden. <sup>8</sup>Trium, Analysis Online GmbH, Hohenlindenerstraße 1, 81677 München, Germany. <sup>9</sup>Wellcome Trust Sanger Institute, Genome Campus, Hinxton CB10 1SA, Cambridge.